Topogenesis of heterologously expressed fragments of the human Y4 GPCR

Marino, J ; Geertsma, E R ; Zerbe, O

Abstract: Fragments of large membrane proteins have the potential to facilitate structural analysis by NMR, but their folding state remains a concern. Here we determined the quality of folding upon heterologous expression for a series of N- or C-terminally truncated fragments of the human Y4 G-protein coupled receptor, amounting to six different complementation pairs. As the individual fragments lack a specific function that could be used to ascertain proper folding, we instead assessed folding on a basic level by studying their membrane topology and by comparing it to well-established structural models of GPCRs. The topology of the fragments was determined using a reporter assay based on C-terminal green fluorescent protein- or alkaline phosphatase-fusions. N-terminal fusions to Lep or Mistic were used if a periplasmic orientation of the N-terminus of the fragments was expected based on predictions. Fragments fused to Mistic expressed at comparably high levels, whereas Lep fusions were produced to a much lower extent. Though none of the fragments exclusively adopted one orientation, often the correct topology predominated. In addition, systematic analysis of the fragment series suggested that the C-terminal half of the Y4 receptor is more important for adopting the correct topology than the N-terminal part. Using the detergent dodecylphosphocholine, selected fragments were solubilized from the membrane and proved sufficiently stable to allow purification. Finally, as a first step toward reconstituting a functional receptor from two fragments, we observed a physical interaction between complementing fragments pairs upon co-expression.

DOI: https://doi.org/10.1016/j.bbamem.2012.07.023

Topogenesis of heterologously expressed fragments of the human Y4 GPCR

Jacopo Marino[1], Eric R. Geertsma[2]*, Oliver Zerbe[1]*

[1]Institute of Organic Chemistry, University of Zurich, Winterthurerstrasse 190, CH 8057, Zurich

[2]Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH 8057, Zurich

*Corresponding authors:

Oliver Zerbe
Institute of Organic Chemistry
University of Zurich
Winterthurerstrasse 190
CH-8057 Zurich, Switzerland
phone: +41-44-6354263
fax: +41-44-6356882
email: oliver.zerbe@oci.uzh.ch

Eric R. Geertsma
Department of Biochemistry
University of Zurich
Winterthurerstrasse 190
CH-8057 Zurich, Switzerland
phone: +41-44-6355546
fax: +41-44-6356834
email: e.geertsma@bioc.uzh.ch

Abstract

Fragments of large membrane proteins have the potential to facilitate structural analysis by NMR, but their folding state remains a concern. Here we determined the quality of folding upon heterologous expression for a series of N- or C-terminally truncated fragments of the human Y4 GPCR, amounting to six different complementation pairs. As the individual fragments lack a specific function that could be used to ascertain proper folding, we instead assessed folding on a basic level by studying their membrane topology and by comparing it to well-established structural models of GPCRs. The topology of the fragments was determined using a reporter assay based on C-terminal GFP- or PhoA-fusions. N-terminal fusions to Lep or Mistic were used if a periplasmic orientation of the N-terminus of the fragments was expected based on predictions. Fragments fused to Mistic expressed at comparably high levels, whereas Lep fusions were produced to a much lower extent. Though none of the fragments exclusively adopted one orientation, often the correct topology predominated. In addition, systematic analysis of the fragment series suggested that the C-terminal half of the Y4 receptor is more important for adopting the correct topology than the N-terminal part. Using the detergent DPC, selected fragments were solubilized from the membrane and proved sufficiently stable to allow purification. Finally, as a first step towards reconstituting a functional receptor from two fragments, we observed a physical interaction between complementing fragments pairs upon co-expression.

Keywords: GPCR, membrane protein folding, Y-receptors, topology screening, Mistic, Lep, GFP, PhoA, heterologous overexpression.

1. Introduction

G-protein coupled receptors are of tremendous importance in signal transduction, and hence play a pivotal role in pharmaceutical sciences and serve as targets for many successful drugs. The crystal structure of bovine rhodopsin was published in the year 2000 and since then we have witnessed an increasing number of new GPCR structures [1-7]. All these structures were determined by X-ray crystallography.

NMR spectroscopy has the additional advantage over crystallography of being able to report on dynamics of proteins, but so far no structures of GPCRs were determined using this method. The recently published structures of sensory rhodopsin [8] and proteorhodopsin [9], though not true GPCRs, proved that the method in principle can determine solution structures of large alpha-helical membrane proteins to high resolution, provided that spectra of good quality can be measured. The latter requirement so far has largely impeded the progress in NMR of true GPCRs.

Our group has focused on the study of large fragments of GPCRs, comprising two or more transmembrane (TM) helices [10]. We study these fragments in order to learn about the folding of GPCRs, in particular to learn about the mechanism of TM bundle assembly, possibly from complementing fragments (the so-called split-receptors). The fragments could also facilitate assignments in the entire receptors. The fragments are produced recombinantly in *E. coli* in various flavors of isotope labeling. Recently we could demonstrate that a fragment comprising the first two TM segments, further on referred to as TM1-2, of the yeast GPCR Ste2p adopts a well-defined secondary structure in LPPG micelles [11]. Moreover, NOEs coding for interhelical contacts involving methyl groups could be detected indicating that the Ste2p TM1-2 fragment adopts a

defined tertiary structure. Due to the increase of the overall correlation time for larger protein-detergent complexes, resulting in broader lines, solution NMR experiments for resonance assignments become more difficult. The use of protein fragments allows differential labeling in order to reduce the complexity of spectra and to facilitate resonance assignments for these challenging systems. Unfortunately, the production of well-folded fragments harboring several TM segments is problematic. Expression in *E. coli* as inclusion bodies often results in high yields but requires refolding, which is challenging for alpha-helical membrane proteins. Alternatively, attempts to express the protein in a folded state in the membrane often lead to decreased yields with increasing size [12, 13].

To obtain rapid information about which fragments are likely suitable for more in-depth NMR studies, we herein present an auxiliary method for investigating the folding state of GPCR fragments. Though in general the folding quality of the protein produced is best determined using an assay that measures its specific activity, here this is not possible as the individual fragments do not carry any function. We instead assessed the folding quality by performing a systematic topological survey of the fragments and determining their agreement with a model derived from the high-resolution structure of rhodopsin [14]. The GPCR studied here, the Y4 receptor, presents a human GPCR targeted by the neurohormones of the neuropeptide Y (NPY) family [15-17].

In this work we have systematically truncated one alpha-helix of the Y4 receptor at a time from its N-terminus (here referred to as N-terminal truncations) or from its C-terminus (referred to as C-terminal truncations), producing the six constructs in the series TM2-7, TM3-7 up to TM7 and the six complementing constructs TM1, TM1-2 up to

TM1-6, respectively (Fig. 1). The topology of the fragments was subsequently determined using the well-established GFP/PhoA reporter assay [18]. The fragments were fused to the C-terminus of Lep or Mistic when appropriate, to enable a periplasmic location of the N-terminus of the fragments. Lep contains the two N-terminal TM segments of the *E. coli* Leader peptidase, also known as Signal Peptidase I. Both Lep and Mistic have been reported to facilitate an initial periplasmic orientation of proteins fused to their C-terminus [19] [20, 21]. In addition, Mistic has been suggested to enhance the expression of membrane proteins regardless of the preferred orientation of the N-terminus of the respective protein [22].

The study herein represents our first attempt to investigate the folding state of heterologously produced GPCR fragments in a true membrane environment. Moreover, we compare the use of two N-terminal fusion proteins that have distinct modes of interaction with the *E. coli* inner membrane. We believe that the knowledge derived from this study will contribute to the design of fragments for self-assembly into a functional split receptor with the future aim of developing this system for structural studies by NMR. Accordingly, we demonstrate physical interactions between co-expressed complementing fragment pairs of the Y4 GPCR.

2. Materials and methods

2.1 Cloning of Y4 receptor fragments starting with TM1

The genes coding for Mistic and Lep were amplified by PCR introducing a 5' NdeI and 3' BamHI restriction site; the two DNA fragments were then cloned in a pET21b vector (Novagen) between NdeI and BamHI sites. The gene coding for the human Y4 receptor was amplified by PCR introducing a 5' BamHI site followed by a 3C protease cleavage site (CTGGAAGTTCTGTTTCAGGGTCCG) and a 3' XhoI site. The PCR product was cloned between BamHI and XhoI into the pET21b vectors harboring *mistic* or *lep*. These two vectors were used as PCR templates for all the C-terminal truncations of the Y4 receptor. The Y4 receptor fragments starting with TM1 were designed to have a Mistic-3C protease site or a Lep-3C protease site fused to their N-terminus and green fluorescent protein (GFP) or the alkaline phosphatase (PhoA) to their C-terminus. The terminal residues of Mistic and Lep were E110 and Q86, respectively, and terminal residues of the Y4 receptor fragments are indicated in Suppl. Table 1. Gene fusions coding for Mistic or Lep and Y4 receptor fragments starting with TM1 were cloned using the FX cloning method [23] into derivatives of pBAD24 [24] containing open reading frames for a 3C protease cleavage site, EGFP and a decaHis-tag (pBXC3GH; [23]) or a linker sequence (VPDSYTQVASWTEPFPFC[25], the mature alkaline phosphatase (Δ_{26}-*phoA*) and a decaHis-tag (pBXCAPH; this work). For each construct, 5 µl of the FX cloning reaction was used to transform chemically competent *E. coli* DH5α. All plasmids were sequence verified and used to transform chemically competent *E. coli* SF100 cells [26].

2.2 Cloning of Y4 receptor fragments ending with TM7

The Y4 receptor fragments ending with TM7 were designed to have Mistic or Lep fused to their N-terminus and green fluorescent protein (GFP) or the alkaline phosphatase (PhoA) to their C-terminus. Derivatives of pBXC3GH and pBXCAPH containing open reading frames coding for Mistic or Lep in fusion with the full length Y4 receptor served as templates. These vectors were amplified using 5' phosphorylated primers. The reverse primer overlapped with the sequence coding for the 3C protease site located after Mistic or Lep. The forward primers overlapped with the starting positions in the Y4 receptor indicated in Suppl. Table 1. PCR products were digested with DpnI to remove the template vectors and subsequently gel purified. 50 ng of DNA in a volume of 8 µl was supplemented with 1 µl of 10X T4 ligase buffer and 1 µl of T4 DNA ligase (1U) and subsequently incubated for 1 hour at RT. After heat-inactivating the T4 ligase for 20 min at 65°C, 5 µl of the ligation reaction was used to transform DH5α cells. All constructs were sequence-verified and transformed to *E. coli* SF100 cells.

2.3 Cloning of EmrE and its mutants

The gene encoding for the wild-type EmrE was amplified from genomic DNA of *E. coli* MC1061 by PCR using primers compatible with the FX cloning method. The EmrE mutant R29G/R82S [27] was obtained by overlapping PCRs and cloned in pBXC3GH and pBXCAPH using FX cloning.

2.4 Cloning of fragments pairs for co-expression

Individual genes coding for the complementation pairs TM1-3/TM4-7 and TM1-5/TM6-7 were manipulated in the pBXC3GH vector to: 1) contain a second 3C protease site adjacent to the first site between Mistic and the fragment; 2) remove the 3C protease site between the fragment and GFP; and 3) remove the His-tag for TM4-7 and TM6-7. Subsequently, the constructs were PCR amplified introducing 5' NcoI and a 3' NotI restriction sites (Mistic-TM1-3-GFP-His and Mistic-TM1-5-GFP-His) or 5' NdeI and a 3' PacI restriction sites (Mistic-TM4-7-GFP and Mistic-TM6-7-GFP). The PCR products coding for the complementation pairs were then cloned in one pCDFDuet-1 vector (Novagen) to allow co-expression. In addition, genes coding for Mistic-TM4-7-GFP and Mistic-TM6-7-GFP were also cloned in the pCDFDuet-1 vectors without the gene coding for their complementation partner. These constructs served as negative controls. All vectors were sequence-verified.

2.4 GFP fluorescence determination in whole cells

Overnight cultures were started from single colonies on LB agar plates and grown in 0.7 ml LB supplemented with ampicillin (LB-amp) at 37°C with vigorous shaking in 96 deep-well plates covered with a gas-permeable seal. The overnight cultures were diluted 1:100 in 0.7 ml of LB-amp in a 96 deep-well plate. Cells were grown for 1 hour at 37°C after which the temperature of the incubator was set to 25°C and cultivation continued for 90 minutes after which the cells were induced with 0.2% (w/v) arabinose. The approximate optical density of the culture at this point was OD_{600}= 0.5. Cultivation was continued for four hours and cells were harvested by centrifugation (3000 *g*, 10 min),

washed once with 500 µl of 50 mM KPi, pH 7.5, 150 mM NaCl and resuspended in 300 µl of the same buffer. GFP fluorescence of whole cells was determined on 150 µl aliquots of these cell suspensions in a black 96 well plate (Nunc, Denmark) using an Infinite M1000 plate reader (Tecan, Switzerland) with excitation at 485 nm and emission at 535 nm. OD_{600} values were determined using 50 µl of the cell suspension in a transparent 96 well plate. All the fluorescence values were normalized by OD_{600} and mean values were obtained by performing the experiment at least twice using biological triplicates.

2.5 PhoA activity determination in whole cells

Overnight cultures and cultivation for expression were performed as described above, except that cells were induced with 0.2% (w/v) arabinose already after 1 hr cultivation at 37°C. At this stage the optical density of the culture was approximately OD_{600} = 0.2. Growth was subsequently continued for one hour at 37°C. Ten minutes before harvesting, 1 mM iodoacetamide was added to prevent oxidation of alkaline phosphatase located in the cytoplasm [28]. All buffers used afterwards contained 1 mM iodoacetamide. Cells were washed at 4°C with 300 µl of 1 M Tris-HCl, pH 8.0, and the OD_{600} was measured using 50 µl of the cell suspension in a transparent 96 well plate. The suspension was subsequently pelleted and cells were resuspended in 300 µl of 1 M Tris-HCl, pH 8.0, containing 1.3 mM p-nitrophenyl-phosphate and incubated in the dark at 37°C until a yellow color developed. Next, cells were pelleted (3000 *g*, 10 min, 4°C) and the A_{420} of 150 µl supernatant was determined using an Infinite M1000 plate reader. Mean activity values were obtained from two independent experiments performed using biological triplicates. The absorbance at 420 nm was normalized for the optical density at 600 nm.

2.6 *In gel* fluorescence and solubilisation assay

Protein samples were prepared as described elsewhere [29], except that whole-cell samples corresponding to ~ 2 mg of protein were used. Protein samples were analyzed by 15% SDS-PAGE, and *in gel* GFP fluorescence was visualized with an LAS-3000 imaging system (Fujifilm, Switzerland) and AIDA software (Raytest, Germany). The solubilization assay was done as described in [29] with the exception that 1 mM DTT was added to all the buffers. The detergents dodecylphosphocholine (DPC), 1-myristoyl-2-hydroxy-*sn*-glycero-3-phosphatidyl-glycerol (LMPG), *n*-dodecyl-β-d-maltoside (DDM), octaethylene glycol monododecyl ether ($C_{12}E_8$; all purchased from Anatrace, Canada) were added to a final concentration of 1% (w/v).

2.7 Preparation of membrane vesicles

Overnight cultures of *E. coli* SF100 containing pBXC3GH derivatives for selected Mistic-Y4 fusions were grown at 37°C in Terrific-broth supplemented with 0.1 mg/ml ampicillin (TB-amp). A 1% (v/v) inoculum was used to start growth in four 2.5 liter baffled flasks each containing 1 liter TB-amp. Cells were grown for 1 hour shaking at 37°C after which the temperature was lowered to 25° C over the course of 90 minutes. After induction with 0.2% (w/v) arabinose, cultivation continued for 16 hours. Cells were harvested by centrifugation at 6000 *g* for 15 min at 4°C. The cell pellet was resuspended in ice-cold 50 mM KPi, pH 7.5, 150 mM NaCl to an OD_{600} between 150 and 200. The suspension was supplemented with lysozyme (1 mg/ml), DNase (20 μg/ml) and MgCl (1 mM), homogenized and stirred for 1 hour at 4°C. Cells were disrupted by three passes through a cooled pressure cell (Emulsiflex) at 15 kPsi before protease inhibitor tablets (Roche) were added. Unbroken cells were removed by centrifugation at 15000 *g* for 15

minutes at 4°C. Subsequently, the membrane vesicles in the supernatant were pelleted by centrifugation at 120000 *g* at 4°C for 1 hour. Membrane vesicles were homogenized in 50 mM KPi, pH 7.5, 150 mM NaCl, and 10% glycerol using a Potter tube to a final concentration of 1 gr (wet weight) of membrane vesicles per 2 ml buffer and immediately frozen in liquid nitrogen and stored at -80°C.

2.8 Purification of the Mistic-Y4 fragment-GFP fusions

Membrane vesicles containing Y4 receptor fragments overexpressed as fusions to Mistic (N-terminus) and GFP-His (C-terminus) were solubilized at a concentration of 1 gr vesicles (wet weight) to 10 ml solubilization buffer (50 mM KPi, pH 7.5, 300 mM NaCl, 5% glycerol, 1 mM DTT, 1% (w/v) DPC and 15 mM imidazole), while stirring for 1 hour at 4°C. All subsequent steps were performed at 4°C. Insoluble material was removed by centrifugation at 120000 *g* for 30 minutes. The supernatants were then incubated for 1 hour with Ni-NTA resin (HIS-Select, Sigma, Switzerland; ~0.6 ml resin per gr vesicles) pre-equilibrated in solubilization buffer. The resin was subsequently washed with ten column volumes (CV) 50 mM KPi, pH 7.5, 300 mM NaCl, 5% glycerol, 5 mM DPC and 30 mM imidazole. Proteins were eluted in 0.5 CV fractions with 50 mM KPi, pH 7.5, 300 mM NaCl, 5% glycerol, 5 mM DPC and 300 mM imidazole and subsequently injected into a Superdex 200 (10/300 GL) SEC column (GE), pre-equilibrated with wash buffer devoid of imidazole. Peak fractions were subsequently re-injected.

2.9 Co-expression and co-purification of fragments pairs

Co-expression of the complementing fragment pairs TM1-3/TM4-7 and TM1-5/TM6-7 was done in *E.coli* BL21 (DE3) using the pCDFDuet-1 derivatives described in Section 2.4. Overnight cultures were started from single colonies and grown in 10 ml-LB broth supplemented with 0.1% streptomycin at 37 °C, shaking. The overnight cultures were diluted 1:100 in 600 ml of TB (2 liters/each fragment pair) and cells were grown for 2 hours at 37°C, shaking. Subsequently the temperature was lowered to 25°C over the course of 1 hour, cultures were induced with 0.2 mM IPTG and cultivation was continued for 16 hrs. Cells were harvested and membrane vesicles were prepared as described in Section 2.7. Purification of the fragments pairs was performed as described in Section 2.8, except that all buffers contained 10% glycerol and 150 mM NaCl. In addition, the Ni-NTA resin (~0.5 ml per gr membrane vesicles) was washed more extensively, using four times ten CV of washing buffer (50 mM KPi, pH 7.5, 150 mM NaCl, 10 % glycerol, 5 mM DPC), supplemented with 30, 5, 30 and 5 mM imidazole, respectively. Proteins were finally eluted in 0.5 CV fractions using an elution buffer containing 300 mM imidazole. Expression and purification of the negative control samples was performed identically.

3. Results.

3.1 Design of the fusion proteins.

Structural characterization of proteins by NMR requires high and robust protein expression levels, and should ideally also include the possibility for perdeuteration. As a result, expression of labeled protein is almost exclusively done in *E. coli* where the protein is produced to high levels in inclusion bodies. Subsequent refolding is required to obtain well-folded material. Unfortunately, this robust strategy is often not successful for alpha-helical membrane proteins due to the difficulties in refolding them in a functional form [12, 13]. To some extend the formation of inclusion bodies can be avoided by decreasing the expression rate, e.g., by lowering the cultivation temperature or using a weaker promoter, thus allowing expression in the membrane in a folded state [29].

To determine which fragments of the Y4 receptor were amenable to overexpression in a folded state, we created a set of six truncation mutants from the N-terminus and a set of six truncation mutants from the C-terminus, differing in length by multiples of single alpha-helices (Fig. 1). To allow for correct processing by the membrane protein insertion machinery, for each fragments starting with a residue that was expected to be on the outside of the cell based on a structural model, the N-terminus was fused to Lep or Mistic. Lep adopts a well-characterized N_{out}-C_{out} topology, while for Mistic, a protein from *B. subtilis* that associates tightly with the membrane, the topology is less well defined (*vide infra*). However, Mistic has been suggested to facilitate and enhance expression of both membrane proteins with an N_{out} or N_{in} topology [22] [20, 21]. For this reason, we also fused Mistic to fragments with a proposed cytoplasmic N-terminus. We

observed that without Lep or Mistic fusions, fragments localize in the inclusion bodies when expressed in *E. coli* [30].

Lacking a functional assay to determine the specific activity of the fragments, we used the membrane topology as an initial indicator for a correct folding state. As reporters for a cytoplasmic or periplasmic location of the C-terminus, the well-established topology indicators GFP and PhoA were used [18, 31]. These indicator proteins were fused to the C-terminus of each fragment. Only GFP located in the cytoplasm will become fluorescent as it remains "dark" when expressed in the periplasm. In contrast, only PhoA expressed in the periplasm will mature and become functional while cytoplasmic PhoA remains inactive if cultivation is limited to several hours [28]. In addition to this, GFP is used as a reporter for protein folding as well [18, 29, 32]. Since the GFP signal reports both folding and topology, it potentially complicates the analysis. Nevertheless a good agreement has been found between membrane topology models determined by PhoA/GFP reporter fusions and the corresponding high-resolution 3D structures [33-38]. Possibly, the PhoA fusion protein reports on folding as well to a similar extend as GFP does.

Except for PhoA, all N- and C-terminal fusion proteins were separated from the Y4 receptor fragments by a 3C protease site to allow future analysis of the fragment only. All combinations of Y4 fragments and different N- and C-terminal fusion partners, amounting to 46 different constructs, were expressed in the same *E. coli* strain, SF100. This strain is compatible with both GFP and PhoA assays [26].

3.2 Topology of the Y4 receptor fragments

For each Y4-fragment, we determined the PhoA activity and the GFP fluorescence of the corresponding fusion proteins using biological triplicates (Fig. 2A). Constructs displaying high GFP fluorescence and low PhoA activity are indicative of membrane proteins with a pronounced cytoplasmic localization, whereas high PhoA activity and low GFP fluorescence are indicative of a periplasmic localization of the C-terminus. Mixed or dual topology is characterized by proteins displaying both significant GFP and PhoA signals. To validate our setup, we made use of three proteins for which the location of the C-terminus had been determined previously. LacS is a transport protein with twelve TM segments and a cytoplasmic C-terminus [39]. Lep contains the first two TM segments preceding the periplasmic domain of the membrane-bound protease Leader Peptidase [19]. Next to these membrane proteins with one clear topology, we used EmrE as an example of a protein with a dual topology. More specifically, we used EmrE(R29S/R82S) as this mutant shows a more similar population of both topologies compared to wild-type EmrE. The dual topology of EmrE has been shown to be essential for its activity [27]. We observed high GFP fluorescence for LacS and high PhoA activity for Lep as expected for membrane proteins with a C_{in} or C_{out} topology, respectively. In contrast, EmrE(R29S/R82S) displayed intermediate values (Fig. 2A). Taken together, these results confirmed the predictive value of the GFP/PhoA reporter assay. We subsequently analyzed Mistic as well to determine the preferred location of its C-terminus. Though Mistic has been shown to facilitate the expression of proteins with an N_{out} topology [22], our data indicate that Mistic not fused to any Y4 fragment has its C-terminus located in

the cytoplasm. To normalize data, we arbitrarily set the high GFP fluorescence of Mistic-GFP and the high PhoA activity of the Lep-PhoA fusion protein to 1.

Fusions between Mistic and the Y4 receptor fragments generally displayed higher GFP and PhoA signals than the corresponding Lep-fusions, which is indicative of higher expression levels (Fig. 2A). For both Mistic and Lep fusions, it was in general difficult to unambiguously assign a defined topology, as often the difference in activities of both reporter proteins was small. To correct for differences in the expression levels of the different fusions, we plotted the ratio of the PhoA to GFP signals for each fragment (Fig. 2B). A PhoA/GFP ratio of 1, or a Log(ratio) of 0, in principle corresponds to a 1:1 mixed topology (50% C_{out} and 50% C_{in}). However, we like to emphasize that the PhoA/GFP assay is not sufficiently quantitative to determine topology distributions in such an exact fashion as the scaling of both axis is arbitrary. Fusions between Mistic and Y4 fragments starting with TM1 showed an alternating PhoA/GFP ratio with every extra TM segment (Fig. 2B). Constructs with an odd number of TM segments display a low PhoA/GFP ratio, whereas constructs with an even number of TM segments show a higher PhoA/GFP ratio, in agreement with the expected topology for GPCRs. In addition, for nearly every extra TM segment added, the topology becomes more skewed towards one particular state, indicating that the fraction of well-oriented fragments and thus the overall quality of folding is improved in the longer constructs.

Interestingly, the opposite trend is observed for the fusions between Mistic and N-terminally truncated Y4 fragments. Here, shortening the constructs by progressively removing TM segments from the N-terminus leads to a more clearly defined C_{in}-topology. These observations suggest that the C-terminus rather than the N-terminus of

the protein is important for establishing the correct topology in the entire receptor. Overall, similar trends were observed when using fusions between Lep and Y4 fragments, though the differences in the PhoA/GFP ratios were less pronounced compared to the Mistic-based fusions.

The average log PhoA/GFP ratio for the fusions of Mistic and Y4 fragments starting with TM1 was lower (-0.44) than the average ratio of the corresponding Lep fusions (+0.22). As all constructs were measured under identical conditions, this suggests that either the Lep fusions have a higher tendency for a periplasmic C-terminus, or the Mistic fusions incline more towards a C-terminus located in the cytoplasm.

To conclude, our results indicate that most probably the Y4 fragments studied here exist as mixtures of proteins with correct and incorrect locations of the C-terminus. However, the distribution of the topologies is not random. For the Y4 fragments starting with TM1 a clear trend in the PhoA/GFP ratio can be seen, suggesting an increased preference of the fragments towards the correct topology. Likewise, most Y4 fragments ending with TM7 display comparably low PhoA/GFP ratios in line with a correct location of the C-terminus in the cytoplasm. These data are indicative of a correct topology for a predominant fraction of the material.

3.3 Purification of individual fragments in fusion with Mistic

As the Mistic-fusions showed in general: 1) more pronounced differences in the PhoA/GFP ratio suggesting a more defined topology; and 2) higher absolute PhoA- and GFP-signals suggesting higher expression levels, we selected only Mistic-Y4 fragments for further analysis. Furthermore, we observed that expression of Mistic-Y4 fragments

affected cell growth less than the corresponding Lep-fusions resulting in approximately five-fold higher final cell densities (data not shown).

We selected the Mistic-Y4 fragment complementation pairs TM1-3 plus TM4-7, and TM1-5 plus TM6-7 for further analysis based on their pronounced PhoA/GFP ratios, and comparably high expression levels. The fragments were expressed identically as for the topology assay to assure that the quality of the material was not affected. All fragments contained a C-terminal His-tagged GFP to facilitate detection and purification. As stability of purified membrane proteins depends strongly on the detergent used for solubilization, we screened four different detergents for their efficiency to solubilize these fragments. The detergents DDM and $C_{12}E_8$ are commonly used for crystallization of membrane proteins [40]. The detergents DPC and LMPC were selected since they have proven to be particularly useful for NMR spectroscopy [41-43]. We compared whole cell lysates solubilized with the respective detergents with the supernatant of the same lysate after ultracentrifugation (Fig. 3C). Provided solubilization is successful, the intensities of the GFP signal before and after centrifugation are expected to be equal. In contrast, aggregation or incomplete solubilization will reduce the GFP signal in the supernatant after centrifugation, as the insoluble material will pellet. Using DDM and $C_{12}E_8$, all four selected fragments and the full-length receptor were completely removed from the solubilization mixture, indicating poor stability and/or incomplete solubilization using these detergents. LMPC only solubilized a fraction of the Y4 receptor fragments. Detergent-solubilization tests indicated that only DPC was able to quantitatively recover the fragments (Fig. 3C). DPC is generally known for its high efficiency in solubilizing membrane proteins [12, 40, 44], but in addition has been shown to solubilize even

misfolded membrane proteins [29]. In contrast, DDM, a mild detergent that tends to selectively solubilize well-folded membrane proteins [29], did not solubilize the fragments. Though this preference for DPC could be related to the folding state, there are many examples of membrane proteins with distinct preferences for certain detergents, and this property could be a peculiarity of these fragments as well.

Using DPC we were able to purify the four fragments and the full-length receptor to ~90% purity by IMAC (Fig. 3A). We obtained approximately 3-5 mg of each Y4 receptor fragment per liter of culture. For all the fragments two major bands were observed with a difference in molecular weight of 10-15 kDa. Only the lower band was fluorescent (Fig. 3B). This dual migration behavior has been observed before and has been correlated to a well-folded fraction, represented by the fluorescent band with a higher electrophoretic mobility, and a fraction of aggregated material, represented by the non-fluorescent band with lower electrophoretic mobility [29]. However, here the signal of the GFP reports not only on the folding quality of the fragments, but on their membrane topology as well. Like aggregated protein, protein inserted in the opposite topology is likely to be misfolded and not expected to solubilize efficiently nor show similar migration properties as well-folded protein. To verify whether a fraction of the protein is indeed not in a well-folded state, or whether the non-fluorescent GFP-fusions represented a fraction folded but misoriented fusion proteins, we analyzed the Y4 receptor fragments by size exclusion chromatography (SEC). Analysis of the samples by SEC revealed identical migration for both fluorescent and non-fluorescent GFP-fusion protein suggesting that both species do have a similar stable folding state (Fig 3B). Only a fraction of the material was aggregated and eluting in the void fraction (Suppl. Fig. 1). The remaining protein

migrated in a broad oligo-disperse peak of which the peak fraction proved quite stable. Upon reinjection of this peak fraction, one relatively broad single peak was obtained (Fig. 3D).

Further analysis of the Y4-fragments by NMR would require removal of the fusion partners as these would complicate the analysis of the NMR spectra and introduce undesired heterogeneity. For this purpose, both Mistic and GFP were connected to the Y4 fragments by a 3C protease cleavage site. Digestion of the purified fusion proteins with the 3C protease readily cleaved of the GFP moiety, but was very inefficient in cleaving the protease site between Mistic and the Y4-fragments (data not shown). Addition of a second 3C protease site between Mistic and the Y4 fragments resulted in a virtually complete digest (Suppl. Fig. 2).

3.4 Co-purification of complementing Y4-fragment pairs.

To simplify resonance assignments in the NMR analysis, the Y4 GPCR fragments should be produced separately to allow differential isotopic labeling. This can be done by inducing proteins production at different stages of cells growth [45]. To determine whether the selected complementing Y4-fragments would in principle be able to stably interact, we co-expressed the complementary pairs TM1-3 plus TM4-7 and TM1-5 plus TM6-7. All fragments were expressed as N-terminal Mistic and C-terminal GFP fusions. Only fragment TM1-3 and TM1-5 additionally contained a C-terminal decaHis-tag.

For both complementation pairs, co-expression of both subunits was succesfull based on in gel detection of GFP fluorescence (Fig. 4, lane SOL). The solubilized fragments were subsequently purified by IMAC. As only one of the subunits contained a His-tag, specific

co-purification of the complementing fragment would be a strong indication for significant affinity of both fragments for each other. To ensure specific co-purification, the IMAC column was washed extensively. For both complementation pairs, we observed co-elution of the untagged Y4 fragment with the His-tagged fragment (Fig. 4, lane Elu). Notably, for each complementation pair the GFP intensity ratio in the solubilization fraction and the elution fractions was clearly changed towards a decreased signal for the untagged Y4 fragment. This was most prominent for the TM1-3 plus TM4-7 pair. The decrease of the GFP signal of the untagged subunits suggested that only a part of the fragments is engaged in a stable interaction. This could be related to our previous observation that the heterologously expressed Y4 fragments are not uniform in the sense that multiple topologies coexist, though the correct topology predominates. Furthermore, the affinity of the complementing fragment might not suffice for complete retention of the untagged fragment. Moreover, a control experiment in which only the untagged subunit was purified, did not reveal its presence in the elution fractions (Fig. 4, right side). This suggests that the presence of the untagged fragments in the elution fractions must indeed result from tight interactions of complementary fragment pairs.

4. Discussion.

Structural characterization of membrane proteins by NMR is still in its infancies despite significant recent progress (reviewed in [46, 47]). In order to enable studies of larger membrane proteins, we are presently exploring alternative approaches using large fragments of the human Y4 GPCR as a model system for two purposes: first, we intend to obtain structural data on these fragments to learn about the assembly of the TM bundle and to obtain chemical shift assignments that may be useful when studying the entire receptor; second, we plan to split the entire receptor into two large fragments that will be produced separately to allow individual isotope-labeling schemes. Upon subsequent mixing the fragments should reconstitute into a fully functional receptor, for which resonance assignments are facilitated because only part of the protein is labeled. Complementation approaches have been pursued before for rhodopsin [48], the β-adrenergic [49] and the Ste2p receptors [50]. These studies showed that functional receptors could be reconstituted upon co-expression of fragments in eukaryotic cells. To the best of our knowledge, the production of interacting GPCR fragments in prokaryotic cells has not been reported so far.

Herein we constructed an extensive set of expression vectors for Y4 receptor fragments truncated from the N- or C-terminus thereby covering each possible combination of complementary fragments (Fig.1). To ensure an appropriate location of the N-terminus and to enhance protein expression levels, we fused the N-termini of the fragments to either Lep or Mistic. In addition, we determined the quality of the fragments expressed. Though protein quality is preferably assessed by determining specific activities, this approach is here not feasible as the fragments do not posses any function. As an

alternative, here we rapidly probed the folding quality of the fragments on a basic level by analyzing their topology within the context of a biological membrane.

For both Lep- and Mistic-fusions starting with TM1, we observed an increase in the PhoA/GFP ratio with fragments containing an even number of TM segments, and a decrease in the ratio for fragments with an odd number, in agreement with their predicted periplasmic and cytoplasmic orientations (Fig. 2B). Despite the appropriate trends of the PhoA/GFP ratios, their absolute values mostly span a range in between those of proteins with a clear C_{in} or C_{out} topology such as LacS or Lep. This suggests that for most fragments both properly oriented and misoriented species co-exist. Nevertheless, a clear preference for the correct topology could be observed, providing an early indication that a significant fraction of the material was correctly folded in the membrane.

Next to this, we observed that the Y4 fragments adopted a less mixed topology if the C-terminal TM segments represented a larger fraction of the Y4 fragment (Fig. 2B). This suggests that the C-terminal part of the Y4 GPCR has a larger influence on adopting the predicted topology than the N-terminus. Hessa *et al.* established a biological scale for determining the propensity of a TM segment to insert in a biological membrane based on its amino acid sequence [51]. This scale considers the TM segments as isolated helices and thereby explicitly neglects the contributions from interhelical interactions. In addition, the preferred orientation of the TM segments in the membrane correlates strongly to the amount of positive charges in the connecting loops (positive-inside rule, [52]). The distribution of charges in the loops of the Y4 GPCR is conform to the positive-inside rule (Fig. 5A). Importantly, the calculated propensities for the TM segments of the Y4 GPCR using the biological partitioning scale are in agreement with our observations

(Fig. 5B). The N-terminal TM segments are predicted to not fully integrate into the membrane, indicating that they require interhelical interactions for proper membrane insertion. Their weak propensity to insert may explain the multiple topologies of TM1 up to TM1-3 as derived from the GFP/PhoA ratios. Interestingly, our recent studies of the Y4 TM1-TM2 fragment in a model membrane system (LPPG micelles) using NMR techniques also revealed that secondary structure in the helices is not stably formed, and that both helices do not fully integrate into the micelles [53]. In contrast, TM segments in the C-terminal half of the receptor have a larger propensity to insert into the membrane and thereby may help in stabilizing the N-terminal TM segments in a hydrophobic environment. As a result, the mixture of topologies is skewed towards their native orientation resulting in PhoA/GFP ratios reflecting more unique topologies.

Though the use of Mistic- and Lep-fusions to the N-terminus of fragments with a predicted N_{out} topology provided qualitatively similar results, a few important differences were noted. First, only the Mistic-fusions expressed at levels compatible with economical production of isotope-labeled protein (3-5 mg protein after IMAC/liter culture), in agreement with previous observations [22, 54]. Second, the average PhoA/GFP ratios for the Mistic- and Lep-fusions differed greatly: Lep-fusions displayed higher average PhoA/GFP ratio than Mistic-fusions (Figure 2A). As the PhoA and GFP signals cannot be directly compared, it is not possible to specifically attribute this difference to a higher average PhoA activity for Lep-fusions or a higher average GFP-activity for Mistic-fusions. It is however in line with the preferred topologies determined for the fusion proteins alone (C_{out} for Lep, C_{in} for Mistic).

Based on the high expression levels and on the PhoA/GFP ratios that suggested a significant fraction of protein with a proper folding state, we selected the Mistic- and GFP fusions of the complementation pairs TM1-3 plus TM4-7, and TM1-5 plus TM6-7 for further analysis. These fragments were expressed well in *E.coli* and could be solubilized efficiently in DPC detergent. The purified proteins proved to be stable in detergent. Moreover, the fusion proteins could be removed using the 3C protease, although in case of Mistic two 3C sites were required for proper processing.

Though the purified Y4 GPCR fragments showed good stability upon reinjection of the main SEC peak, the mixed topologies, the requirement for DPC, the dual migration of the GFP-fusions and the broad peak during the initial analysis by SEC raised concerns about the folding state of the protein produced. To determine whether the selected complementing Y4-fragments are in principle able to stably interact, the complementation pairs TM1-3 plus TM4-7 and TM1-5 plus TM6-7 were co-expressed. Upon purification, we observed specific co-elution of the untagged TM4-7 and TM6-7 with their respective His-tagged complementation partners TM1-3 and TM1-5. This suggests that at least a significant fraction of these Y4 GPCR fragments is in a state that allows a strong physical interaction with its complementation partner. Our topological study suggests that the fragment starting with the N terminus is better folded the longer it is, and hence the TM1-5 fragment is expected to be preferable over the TM1-3 fragment. Our co-purification data indicate that this is indeed the case. We conclude with the statement that our study suggests that producing complementary fragments presents a promising route to obtain a reconstituted receptor that can be a useful system for simplifying its spectroscopic characterization by NMR.

Acknowledgements

We acknowledge financial support by the Swiss National Science Foundation (grant No 31003A_124649). For the kind gift of plasmids and strains we are indebted to Prof Senyon Choe (*mistic*), Prof. Gunnar von Heijne and Dr. IngMarie Nilsson (pRD8 containing *lep*), Prof. J.S. Lolkema and Ramon ter Horst (pLIC1 containing *phoA* and *E. coli* SF100), and Prof. Raimund Dutzler (pBXC3GH). We thank Prof. R. Dutzler for hosting J.M. during this study. We would also like to thank Sandra Steiner for helpful discussions regarding the treatment of the statistical data. E.R.G acknowledges a long-term fellowship from the Human Frontier Science Program.

References:

[1] Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M. and Miyano, M. Crystal structure of rhodopsin: A G protein-coupled receptor. Science 289 (2000) 739-745.

[2] Warne, T., Serrano-Vega, M. J., Baker, J. G., Moukhametzianov, R., Edwards, P. C., Henderson, R., Leslie, A. G., Tate, C. G. and Schertler, G. F. Structure of a beta1-adrenergic G-protein-coupled receptor. Nature 454 (2008) 486-491.

[3] Rasmussen, S. G., Choi, H. J., Rosenbaum, D. M., Kobilka, T. S., Thian, F. S., Edwards, P. C., Burghammer, M., Ratnala, V. R., Sanishvili, R., Fischetti, R. F., Schertler, G. F., Weis, W. I. and Kobilka, B. K. Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. Nature 450 (2007) 383-387.

[4] Jaakola, V. P., Griffith, M. T., Hanson, M. A., Cherezov, V., Chien, E. Y., Lane, J. R., Ijzerman, A. P. and Stevens, R. C. The 2.6 Å crystal structure of a human A2A adenosine receptor bound to an antagonist. Science 322 (2008) 1211-1217.

[5] Chien, E. Y., Liu, W., Zhao, Q., Katritch, V., Han, G. W., Hanson, M. A., Shi, L., Newman, A. H., Javitch, J. A., Cherezov, V. and Stevens, R. C. Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. Science 330 (2010) 1091-1095.

[6] Wu, B., Chien, E. Y., Mol, C. D., Fenalti, G., Liu, W., Katritch, V., Abagyan, R., Brooun, A., Wells, P., Bi, F. C., Hamel, D. J., Kuhn, P., Handel, T. M., Cherezov, V. and Stevens, R. C. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. Science 330 (2010) 1066-1071.

[7] Shimamura, T., Shiroishi, M., Weyand, S., Tsujimoto, H., Winter, G., Katritch, V., Abagyan, R., Cherezov, V., Liu, W., Han, G. W., Kobayashi, T., Stevens, R. C. and Iwata, S. Structure of the human histamine H1 receptor complex with doxepin. Nature 475 (2011) 65-70.

[8] Gautier, A., Mott, H. R., Bostock, M. J., Kirkpatrick, J. P. and Nietlispach, D. Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy. Nat. Struct. Mol. Biol. 17 (2010) 768-774.

[9] Reckel, S., Gottstein, D., Stehle, J., Löhr, F., Verhoefen, M. K., Takeda, M., Silvers, R., Kainosho, M., Glaubitz, C., Wachtveitl, J., Bernhard, F., Schwalbe, H., Güntert, P. and Dötsch, V. Solution NMR Structure of Proteorhodopsin. Angew. Chem. Int. Ed. Engl. 50 (2011) 11942-11946.

[10] Neumoin, A., Arshava, B., Becker, J., Zerbe, O. and Naider, F. NMR studies in dodecylphosphocholine of a fragment containing the seventh transmembrane helix of a G-protein-coupled receptor from Saccharomyces cerevisiae. Biophys. J. 93 (2007) 467-482.

[11] Neumoin, A., Cohen, L. S., Arshava, B., Tantry, S., Becker, J. M., Zerbe, O. and Naider, F. Structure of a double transmembrane fragment of a G-protein-coupled receptor in micelles. Biophys. J. 96 (2009) 3187-3196.

[12] White, M. A., Clark, K. M., Grayhack, E. J. and Dumont, M. E. Characteristics affecting expression and solubilization of yeast membrane proteins. J. Mol. Biol. 365 (2007) 621-636.

[13] Dobrovetsky, E., Lu, M. L., Andorn-Broza, R., Khutoreskaya, G., Bray, J. E., Savchenko, A., Arrowsmith, C. H., Edwards, A. M. and Koth, C. M. High-throughput production of prokaryotic membrane proteins. J. Struct. Funct. Genomics 6 (2005) 33-50.

[14] Bissantz, C., Logean, A. and Rognan, D. High-throughput modeling of human G-protein coupled receptors: amino acid sequence alignment, three-dimensional model building, and receptor library screening. J. Chem. Inf. Comput. Sci. 44 (2004) 1162-1176.

[15] Turton, M. D., O'Shea, D. and Bloom, S. R. (1997) in *Neuropeptide Y and drug development* (Grundemar, L. and Bloom, S. R., Eds.), pp. 15-39, Acedemic Press, New York.

[16] Larhammar, D., Wraith, A., Berglund, M. M., Holmberg, S. K. and Lundell, I. Origins of the many NPY-family receptors in mammals. Peptides 22 (2001) 295-307.

[17] Balasubramaniam, A. Clinical potentials of neuropeptide Y family of hormones. Am. J. Surg. 183 (2002) 430-434.

[18] Drew, D., Sjostrand, D., Nilsson, J., Urbig, T., Chin, C. N., de Gier, J. W. and von Heijne, G. Rapid topology mapping of Escherichia coli inner-membrane proteins by prediction and PhoA/GFP fusion analysis. Proc. Natl. Acad. Sci. U S A 99 (2002) 2690-2695.

[19] Bilgin, N., Lee, J. I., Zhu, H. Y., Dalbey, R. and von Heijne, G. Mapping of catalytically important domains in Escherichia coli leader peptidase. EMBO J. 9 (1990) 2717-2722.

[20] Nekrasova, O. V., Wulfson, A. N., Tikhonov, R. V., Yakimov, S. A., Simonova, T. N., Tagvey, A. I., Dolgikh, D. A., Ostrovsky, M. A. and Kirpichnikov, M. P. A new hybrid protein for production of recombinant bacteriorhodopsin in Escherichia coli. J. Biotechnol. 147 (2010) 145-150.

[21] Blain, K. Y., Kwiatkowski, W. and Choe, S. The functionally active Mistic-fused histidine kinase receptor, EnvZ. Biochemistry 49 (2010) 9089-9095.

[22] Roosild, T. P., Greenwald, J., Vega, M., Castronovo, S., Riek, R. and Choe, S. NMR structure of Mistic, a membrane-integrating protein for membrane protein expression. Science 307 (2005) 1317-1321.

[23] Geertsma, E. R. and Dutzler, R. A versatile and efficient high-throughput cloning tool for structural biology. Biochemistry 50 (2011) 3272-3278.

[24] Guzman, L. M., Belin, D., Carson, M. J. and Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J. Bacteriol. 177 (1995) 4121-4130.

[25] Rapp, M., Drew, D., Daley, D. O., Nilsson, J., Carvalho, T., Melen, K., De Gier, J. W. and Von Heijne, G. Experimentally based topology models for E. coli inner membrane proteins. Protein Sci. 13 (2004) 937-945.

[26] Ter Horst, R. and Lolkema, J. S. Rapid screening of membrane topology of secondary transport proteins. Biochim. Biophys. Acta 1798 (2010) 672-680.

[27] Rapp, M., Seppala, S., Granseth, E. and von Heijne, G. Emulating membrane protein evolution by rational design. Science 315 (2007) 1282-1284.

[28] Derman, A. I. and Beckwith, J. Escherichia coli alkaline phosphatase localized to the cytoplasm slowly acquires enzymatic activity in cells whose growth has been suspended: a caution for gene fusion studies. J. Bacteriol. 177 (1995) 3764-3770.

[29] Geertsma, E. R., Groeneveld, M., Slotboom, D. J. and Poolman, B. Quality control of overexpressed membrane proteins. Proc. Natl. Acad. Sci. U S A 105 (2008) 5722-5727.

[30] Kocherla, H., Marino, J., Shao, X., Graf, J., Zou, C. and Zerbe, O. Biosynthesis and spectroscopic characterization of 2-TM fragments encompassing the sequence of a human GPCR, the Y4 receptor. ChemBioChem 13 (2012) 818-828.

[31] Daley, D. O., Rapp, M., Granseth, E., Melen, K., Drew, D. and von Heijne, G. Global topology analysis of the Escherichia coli inner membrane proteome. Science 308 (2005) 1321-1323.

[32] Waldo, G. S., Standish, B. M., Berendzen, J. and Terwilliger, T. C. Rapid protein-folding assay using green fluorescent protein. Nat. Biotechnol. 17 (1999) 691-695.

[33] Cassel, M., Seppala, S. and von Heijne, G. Confronting fusion protein-based membrane protein topology mapping with reality: the Escherichia coli ClcA H+/Cl- exchange transporter. J. Mol. Biol. 381 (2008) 860-866.

[34] Calamia, J. and Manoil, C. lac permease of Escherichia coli: topology and sequence elements promoting membrane insertion. Proc. Natl. Acad. Sci. U S A 87 (1990) 4937-4941.

[35] Boyd, D., Manoil, C. and Beckwith, J. Determinants of membrane protein topology. Proc. Natl. Acad. Sci. U S A 84 (1987) 8525-8529.

[36] Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H. R. and Iwata, S. Structure and mechanism of the lactose permease of Escherichia coli. Science 301 (2003) 610-615.

[37] Dassa, E. and Muir, S. Membrane topology of MalG, an inner membrane protein from the maltose transport system of Escherichia coli. Mol. Microbiol. 7 (1993) 29-38.

[38] Oldham, M. L., Khare, D., Quiocho, F. A., Davidson, A. L. and Chen, J. Crystal structure of a catalytic intermediate of the maltose transporter. Nature 450 (2007) 515-521.

[39] Gunnewijk, M. G., Postma, P. W. and Poolman, B. Phosphorylation and functional properties of the IIA domain of the lactose transport protein of Streptococcus thermophilus. J. Bacteriol. 181 (1999) 632-641.

[40] Newstead, S., Kim, H., von Heijne, G., Iwata, S. and Drew, D. High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA 104 (2007) 13936-13941.

[41] Warschawski, D. E., Arnold, A. A., Beaugrand, M., Gravel, A., Chartrand, E. and Marcotte, I. Choosing membrane mimetics for NMR structural studies of transmembrane proteins. Biochim. Biophys. Acta 1808 (2011) 1957-1974.

[42] Page, R. C., Moore, J. D., Nguyen, H. B., Sharma, M., Chase, R., Gao, F. P., Mobley, C. K., Sanders, C. R., Ma, L., Sönnichsen, F. D., Lee, S., Howell, S. C., Opella, S. J. and Cross, T. A. Comprehensive evaluation of solution nuclear

magnetic resonance spectroscopy sample preparation for helical integral membrane proteins. J. Struct. Funct. Gen. 7 (2006) 51-64.

[43] Krüger-Koplin, R. D., Sorgen, P. L., Krüger-Koplin, S. T., Rivera-Torres, I. O., Cahill, S. M., Hicks, D. B., Grinius, L., Krulwich, T. A. and Girvin, M. E. An evaluation of detergents for NMR structural studies of membrane proteins. J. Biomol. NMR 28 (2004) 43-57.

[44] Eshaghi, S., Hedren, M., Nasser, M. I., Hammarberg, T., Thornell, A. and Nordlund, P. An efficient strategy for high-throughput expression screening of recombinant integral membrane proteins. Protein Sci. 14 (2005) 676-683.

[45] Zuger, S. and Iwai, H. Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies. Nat. Biotechnol. 23 (2005) 736-740.

[46] Kim, H. J., Howell, S. C., Van Horn, W. D., Jeon, Y. H. and Sanders, C. R. Recent Advances in the Application of Solution NMR Spectroscopy to Multi-Span Integral Membrane Proteins. Prog. Nucl. Magn. Reson. Spectrosc. 55 (2009) 335-360.

[47] Kang, C. and Li, Q. Solution NMR study of integral membrane proteins. Curr. Opin. Chem. Biol. 15 (2011) 560-569.

[48] Ridge, K. D., Lee, S. S. and Yao, L. L. In vivo assembly of rhodopsin from expressed polypeptide fragments. Proc. Natl. Acad. Sci. U S A 92 (1995) 3204-3208.

[49] Kobilka, B. K., Kobilka, T. S., Daniel, K., Regan, J. W., Caron, M. G. and Lefkowitz, R. J. Chimeric alpha 2-,beta 2-adrenergic receptors: delineation of

domains involved in effector coupling and ligand binding specificity. Science 240 (1988) 1310-1316.

[50] Martin, N. P., Leavitt, L. M., Sommers, C. M. and Dumont, M. E. Assembly of G protein-coupled receptors from fragments: identification of functional receptors with discontinuities in each of the loops connecting transmembrane segments. Biochemistry 38 (1999) 682-695.

[51] Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S. H. and von Heijne, G. Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature 433 (2005) 377-381.

[52] von Heijne, G. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. J. Mol. Biol. 225 (1992) 487-494.

[53] Shao, X., Zou, C., Naider, F. and Zerbe, O. Comparison of fragments comprising the first two helices of the human Y4 and the yeast Ste2p G protein coupled receptors, submitted.

[54] Deniaud, A., Bernaudat, F., Frelet-Barrand, A., Juillan-Binard, C., Vernet, T., Rolland, N. and Pebay-Peyroula, E. Expression of a chloroplast ATP/ADP transporter in E. coli membranes: behind the Mistic strategy. Biochim. Biophys. Acta 1808 (2011) 2059-2066.

Figure Captions

Figure 1: Schematic representation of the truncated versions of the human Y4 receptor used in this study. Constructs are divided in C-terminal and N-terminal truncations. Mistic and Lep (blue and gray triangles) were used as fusion proteins to obtain an N_{out} topology for the Y4 receptor fragments where needed and to enhance expression levels of fragments with an N_{in} topology (Mistic only, blue triangles). GFP and PhoA were used as C-terminal reporter proteins for cellular localizations (indicated by a circle).

Figure 2: Membrane topology of Mistic- and Lep-fusion of Y4 GPCR fragments as determined with the PhoA/GFP reporter assay. LacS, Lep and EmrE* represent control membrane proteins with a cytoplasmic, periplasmic or mixed location of their C-terminus, respectively. EmrE* denotes the R29G/R82S double-mutant of EmrE. A) GFP fluorescence (black bars) and PhoA activities (white bars) based on biological triplicates were normalized and plotted in pairs for each individual construct. B) log (PhoA/GFP) values for each individual construct.

Figure 3: Purification of selected fragments of the Y4 GPCR. A) SDS-PAGE analysis of single fragments purified in 5 mM DPC using IMAC and SEC. White and black arrows indicate the non-fluorescent and fluorescent species of the purified constructs, respectively. B) In gel GFP fluorescence of the gel depicted in panel A prior to Coomassie staining. Black arrows indicate the fluorescent target band. C) Differential sedimentation of the Mistic- and GFP-fusions analyzed by in gel fluorescence. Whole cell lysates were disrupted by bead beating and solubilized with 1% DDM, 1% $C_{12}E_8$, 1%

DPC, or 1% LMPC. T and S represent samples taken before (Total protein) and after (Solubilized protein) ultracentrifugation, respectively. D) SEC of reinjected peak fractions from a previous run (Suppl. Fig. 1).

Figure 4: Co-purification of complementing Y4 GPCR fragments pairs. A) Samples taken at different stages of the purification from membranes containing Mistic-TM1-3-GFP-His tag and Mistic-TM4-7-GFP (labelled Test) or Mistic-TM4-7-GFP alone (labelled Negative control). B) Samples taken at different stages of the purification from membranes containing Mistic-TM1-5-GFP-His tag and Mistic-TM6-7-GFP (labelled Test) or Mistic-TM6-7-GFP alone (labelled Negative control). Upper and lower panels indicate the same gel after imaging by in gel GFP fluorescence and Coomassie staining, respectively. Labels refer to marker (M), solubilized membranes (SOL), flow though (FT), first wash (W1), second wash (W2), third wash (W3), fourth wash (W4) and elution (ELU). The position of the fluorescent fraction of the protein is indicated by white and black arrows.

Figure 5: Charge distribution in loops and calculated insertion propensities for individual TM segments of the Y4 GPCR. (A) Net charge of all Y4 GPCR loops. (B) Free energies for transferring entire TM segments into the membrane using a biological scale[56]. Amino acid sequences of the TM helices are indicated in Suppl. Table 2.

Figure 1

Click here to download high resolution image

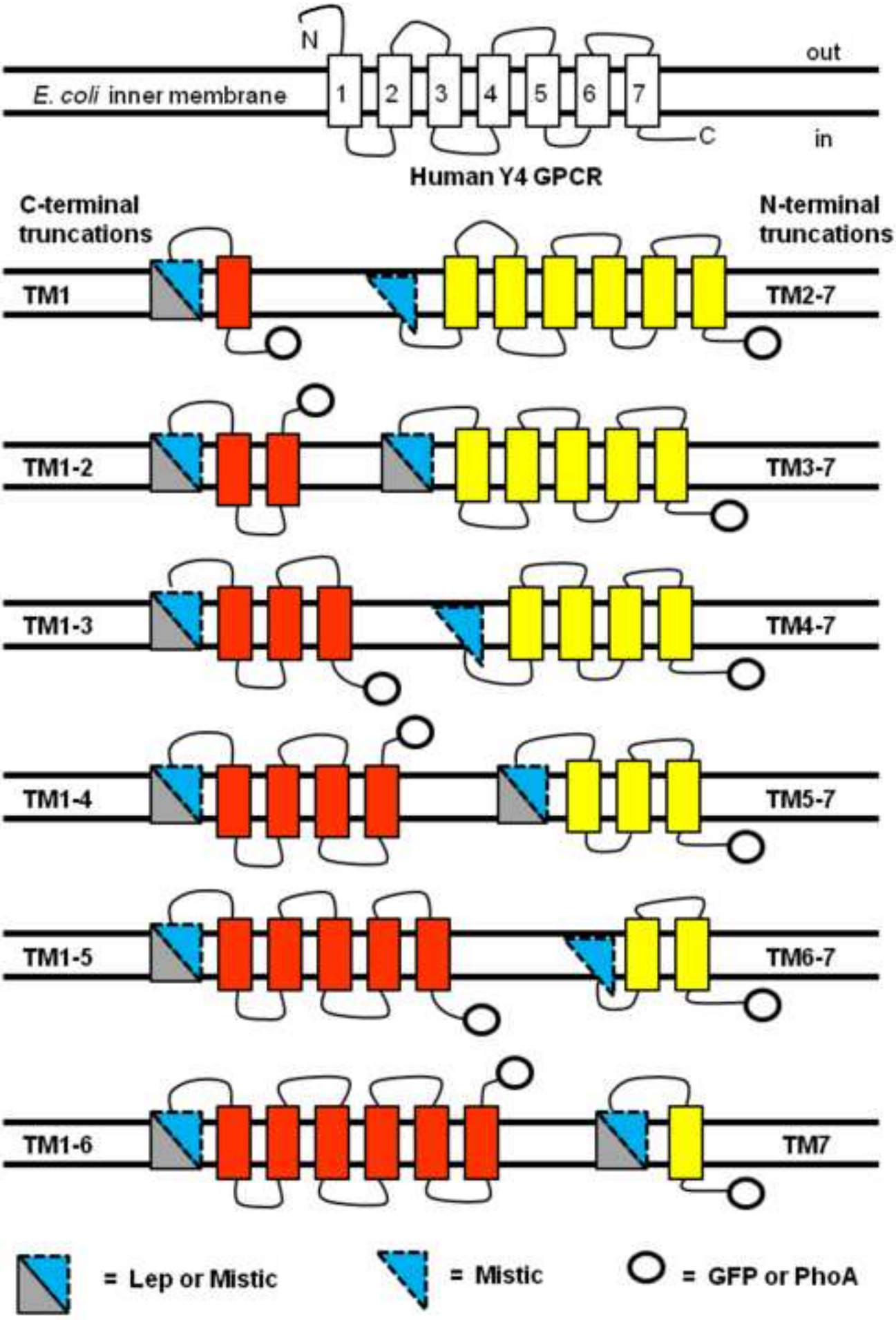


Figure 2

Click here to download high resolution image

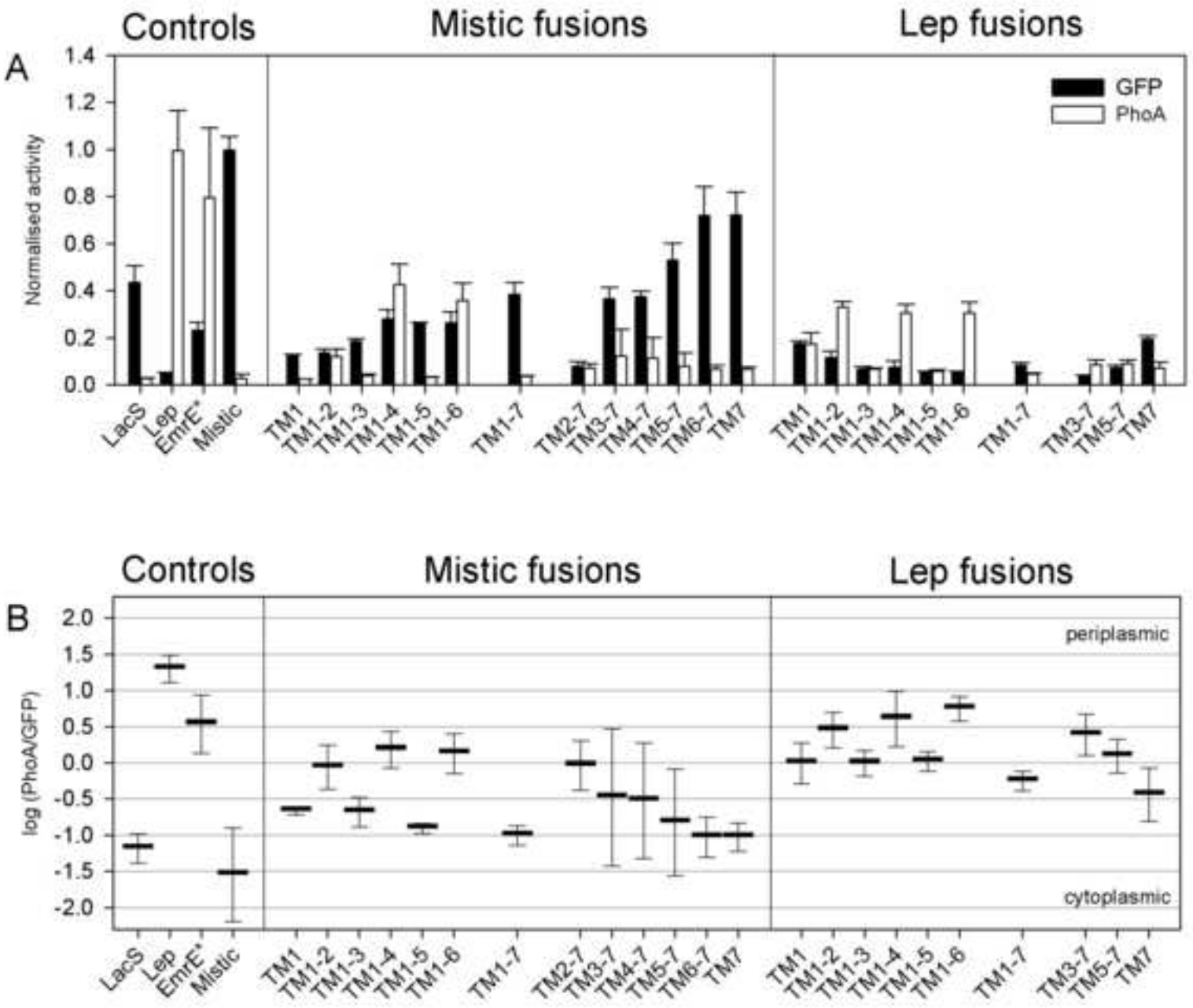


Figure 3

Click here to download high resolution image

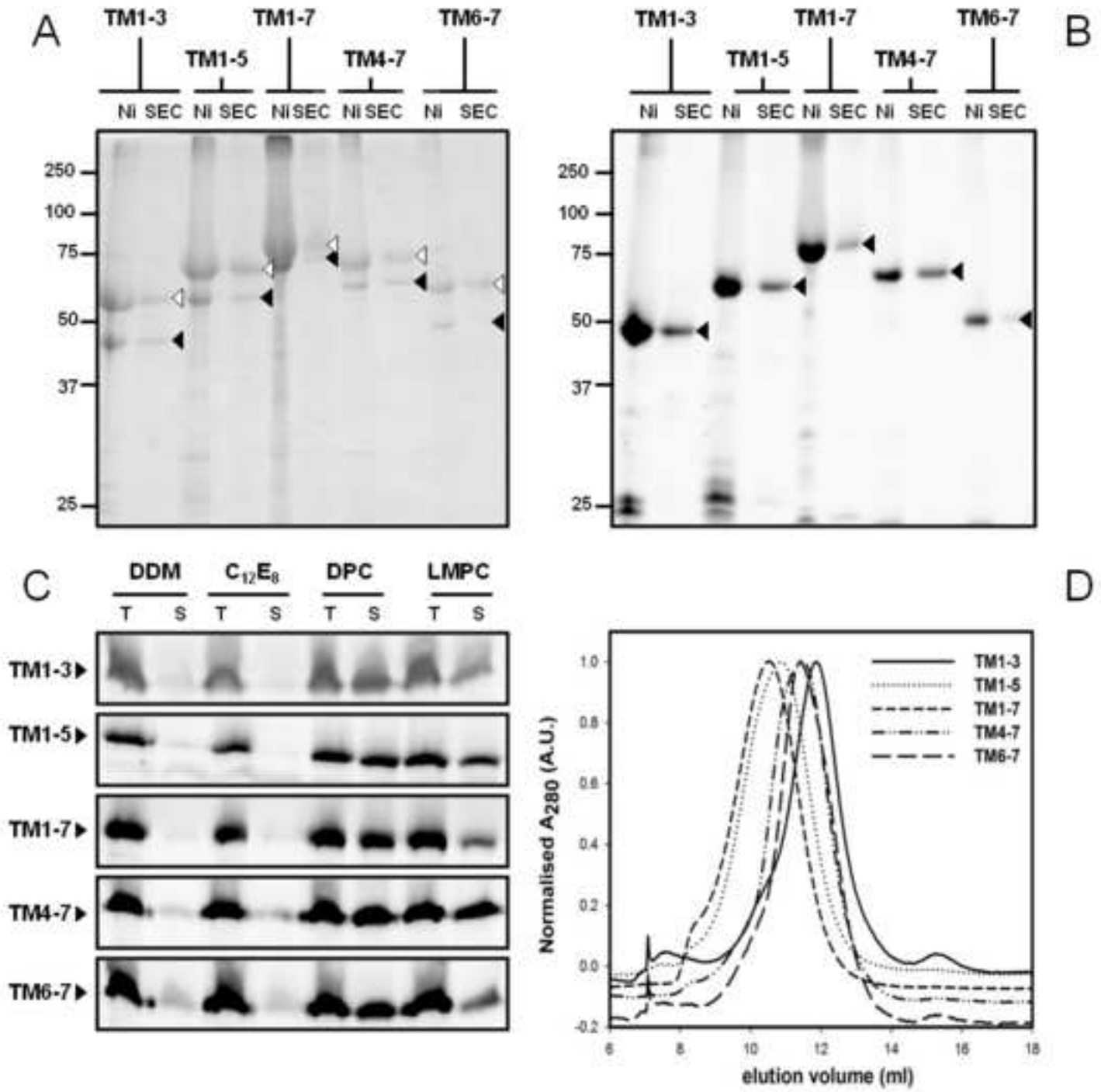


Figure 4

Click here to download high resolution image

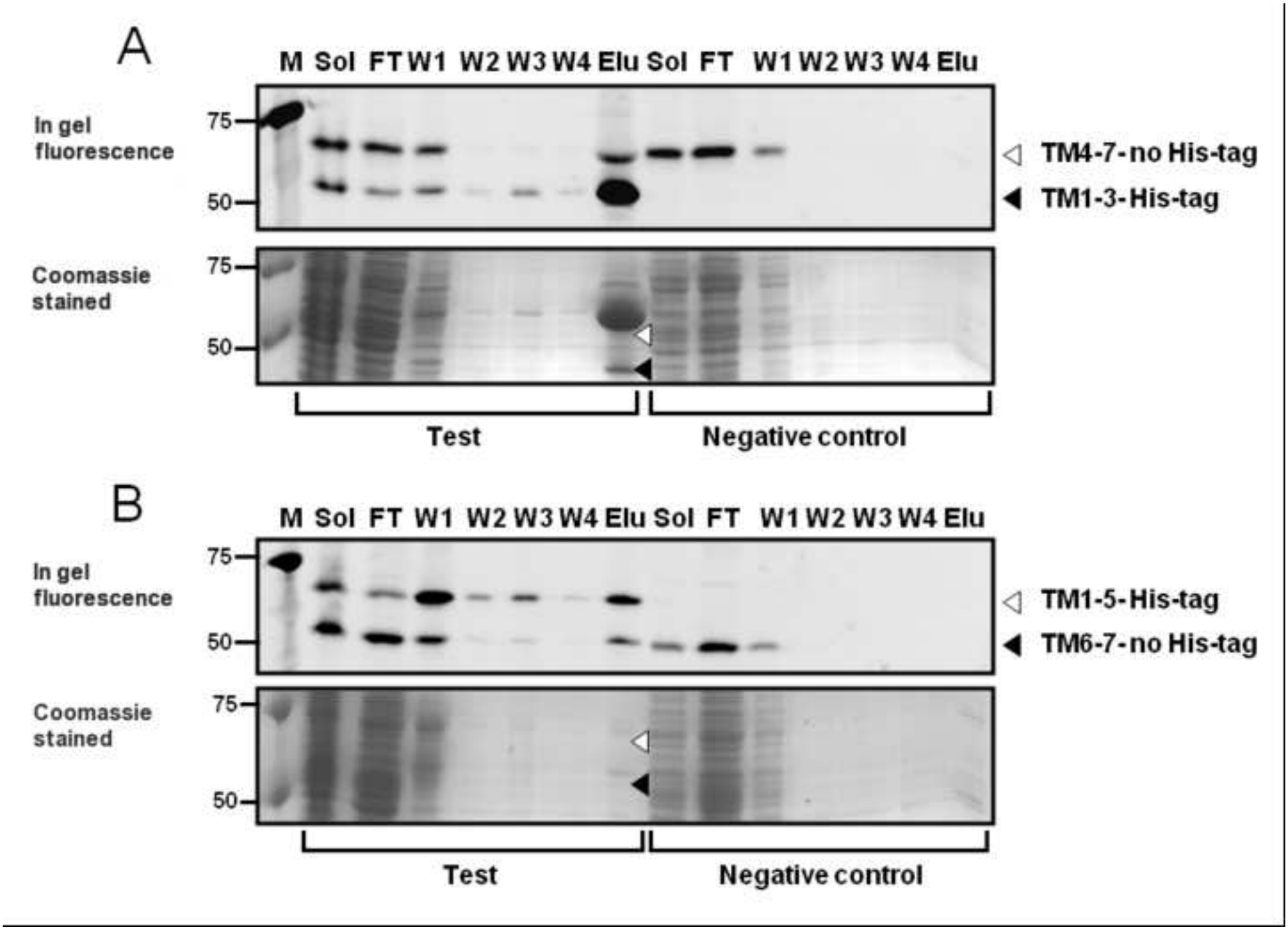
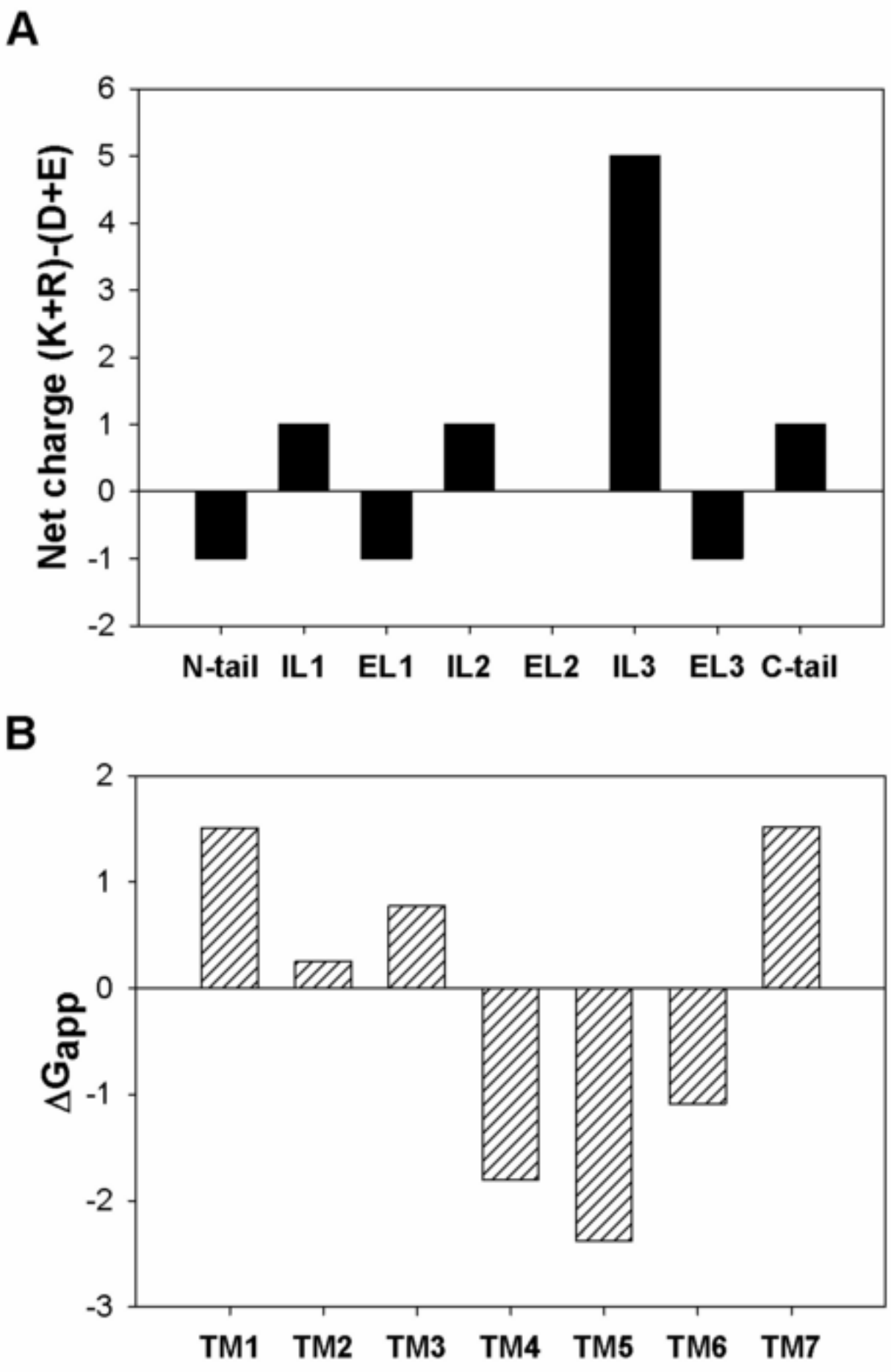


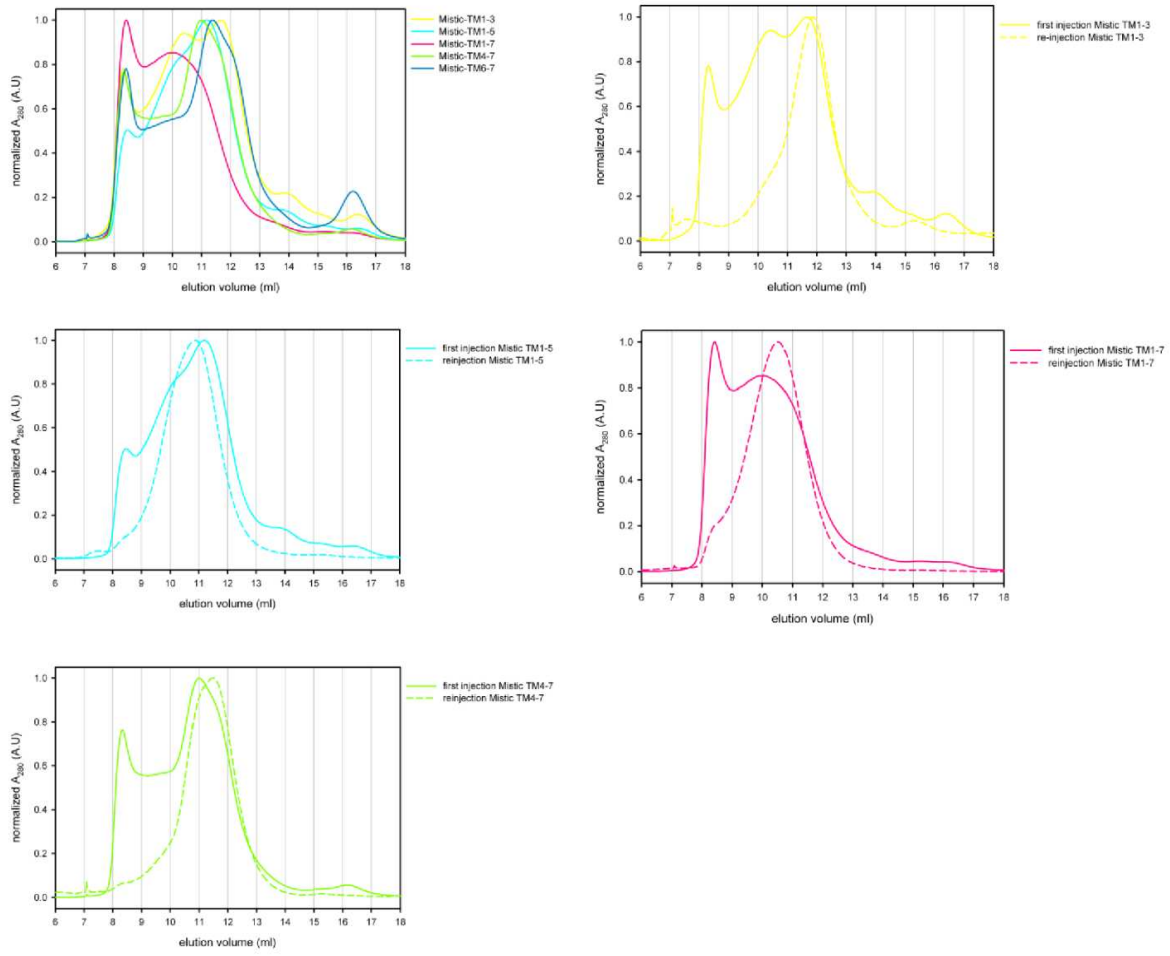
Figure 5
Click here to download high resolution image



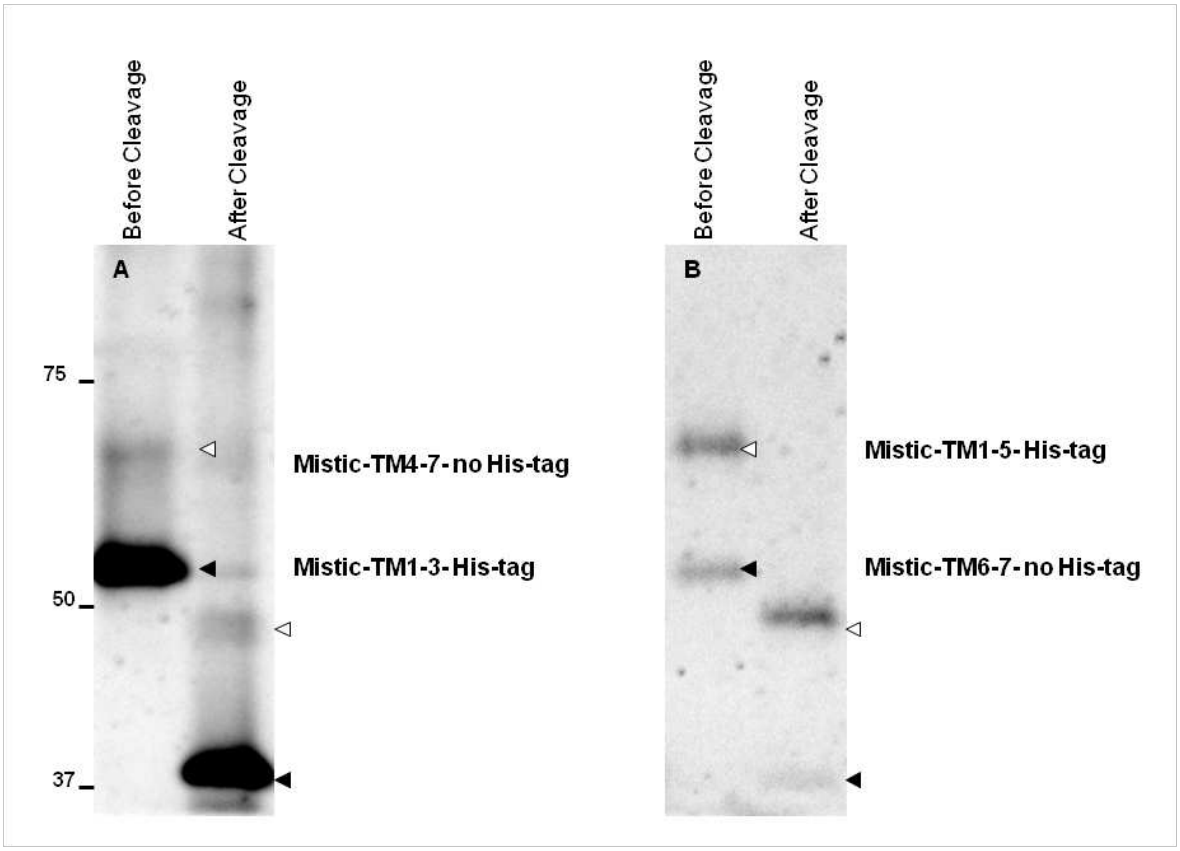
Supplementary material

Topogenesis of heterologously expressed fragments of the human Y4 GPCR

Jacopo Marino[1], Eric R. Geertsma[2]*, Oliver Zerbe[1]*



Suppl. Fig.1: Normalized size exclusion profiles of the constructs TM1-3, TM1-5, TM1-7, TM4-7 and TM6-7. All constructs were N-terminally fused to Mistic and C-terminally to the GFP, purified by IMAC as described in Materials and Methods and subsequently injected in a Superdex 200 (10/300 GL) (GE). Fractions corresponding to the main peak were re-injected (dotted lines).



Suppl. Fig. 2: Enzymatic cleavage of Mistic from the co-purified fragments pairs. Fragments pairs TM13/TM47 and TM15/TM67 were co-expressed and co-purified as described in Materials and Methods. Fractions corresponding to the main elution peaks were used as test for the efficiency of the Mistic removal upon addition of 3C protease. The cleavage was performed at 20 °C for 12 hours in 150mM NaCl, 50mM KPi pH 7.5, 10% glycerol, 1mM DPC. The final concentration of 3C protease was 70 μM.

Fusion protein		N-terminal fragments			C-terminal fragments		
	KDa		KDa	Start-end aa		KDa	Start-end aa
Mistic	12.8	**TM1**	8	M1-K72	**TM2-7**	35	R68-I375
Lep	9.3	**TM1-2**	11.7	M1-Y106	**TM3-7**	30.7	G110-I375
GFP	25	**TM1-3**	17.2	M1-Q155	**TM4-7**	27	N145-I375
		TM1-4	22	M1-K198	**TM5-7**	20.3	K198-I375
		TM1-5	28.4	M1-K251	**TM6-7**	15.4	R239-I375
		TM1-6	33.8	M1-C298	**TM7**	9.7	D289-I375
		TM1-7	42.2	M1-I375			

Suppl. Table 1: Molecular weights and positions at which the N- and C-terminal fragments of the Y4 GPCR have been truncated.

Alpha helix number	Amino acid sequence
TM1	FIVTSYSIETTVVGVLGNLCLMCVTVR
TM2	LLIANLAFSDFLMCLLCQPLTAVYTI
TM3	TLCKMSAFIQCMSVTVSISLVLVAL
TM4	ISQAYLGIVLIWVIACVLSLPLA
TM5	RTIYTTFLLLFQYCLPLGFILVCYARIYRR
TM6	VNVVLVVMVVAFAVLWLPL
TM7	LIFLVCHLLAMASTCV

Suppl. Table 2: Amino acid sequences of the single alpha helices from the Y4 GPCR used as input for calculating the ΔG_{app} for membrane insertion using the ΔG prediction server v1.0 (http://dgpred.cbr.su.se/).